PCT/EP03/02363

Europäisches Patentamt | European Patent Office | Office européen des brevets

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02005141.3

CERTIFIED COPY OF PRIORITY DOCUMENT

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 02005141.3
Demande no:

Anmeldetag:
Date of filing: 07.03.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

DeveloGen Aktiengesellschaft für
entwicklungsbiologische Forschung
Rudolf-Wissell-Strasse 28
37079 Göttingen
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

CG3842 homologous proteins involved in the regulation of energy homeostasis

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12N15/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR

WEICKMANN & WEICKMANN

Patentanwälte
European Patent Attorneys · European Trademark Attorneys

DIPL.-ING. H. WEICKMANN (bis 31.1.01)
DIPL.-ING. F. A. WEICKMANN
DIPL.-CHEM. B. HUBER
DR.-ING. H. LISKA
DIPL.-PHYS. DR. J. PRECHTEL
DIPL.-CHEM. DR. B. BÖHM
DIPL.-CHEM. DR. W. WEISS
DIPL.-PHYS. DR. J. TIESMEYER
DIPL.-PHYS. DR. M. HERZOG
DIPL.-PHYS. B. RUTTENSPERGER
DIPL.-PHYS. DR.-ING. V. JORDAN
DIPL.-CHEM. DR. M. DEY
DIPL.-FORSTW. DR. J. LACHNIT

Unser Zeichen:
27673P EP/WWmh

Anmelder:
DeveloGen Aktiengesellschaft für
entwicklungsbiologische Forschung
Rudolf-Wissell-Straße 28

37079 Göttingen
DEUTSCHLAND

CG3842 homologous proteins involved in the regulation of energy homeostasis

---

Postfach 860 820, 81635 München, Deutschland, Tel. (089) 45563 0, Fax (089) 45563 999. email@weickmann.de

CG3842 homologous proteins involved in the regulation of energy homeostasis

Description

This invention relates to the use of nucleic acid sequences encoding CG3842 or SCAD homologous proteins, and the polypeptides encoded thereby and to the use thereof in the diagnosis, study, prevention, and treatment of diseases and disorders related to body-weight regulation, for example, but not limited to, metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea.

Obesity is one of the most prevalent metabolic disorders in the world. It is still poorly understood human disease that becomes more and more relevant for western society. Obesity is defined as an excess of body fat, frequently resulting in a significant impairment of health. Besides severe risks of illness such as diabetes, hypertension and heart disease, individuals suffering from obesity are often isolated socially. Human obesity is strongly influenced by environmental and genetic factors, whereby the environmental influence is often a hurdle for the identification of (human) obesity genes. Obesity is influenced by genetic, metabolic, biochemical, psychological, and behavioral factors. As such, it is a complex disorder that must be addressed on several fronts to achieve lasting positive clinical outcome. Obese individuals are prone to ailments including: diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancers of the reproductive organs, and sleep apnea.

Obesity is not to be considered as a single disorder but a heterogeneous group of conditions with (potential) multiple causes. Obesity is also characterized by elevated fasting plasma insulin and an exaggerated insulin response to oral glucose intake (Koltermann, 1980, J. Clin. Invest 65:1272-1284) and a clear involvement of obesity in type 2 diabetes mellitus can be confirmed (Kopelman, 2000, Nature 404:635-643).

Even if several candidate genes have been described which are supposed to influence the homeostatic system(s) that regulate body mass/weight, like leptin, VCPI, VCPL, or the peroxisome proliferator-activated receptor-gamma co-activator, the distinct molecular mechanisms and/or molecules influencing obesity or body weight/body mass regulations are not known.

Therefore, the technical problem underlying the present invention was to provide for means and methods for modulating (pathological) metabolic conditions influencing body-weight regulation and/or energy homeostatic circuits. The solution to said technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, the present invention relates to genes with novel functions in body-weight regulation, energy homeostasis, metabolism, and obesity. The present invention discloses a specific gene involved in the regulation of body-weight, energy homeostasis, metabolism, and obesity, and thus in disorders related thereto such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea. The present invention describes the human homolog of the Drosophila CG3842 gene as being involved in those conditions mentioned above.

The term 'GenBank Accession number' relates to NCBI GenBank database entries (Benson et al, Nucleic Acids Res. 28, 2000, 15-18).

BACKGROUND

The acyl-CoA dehydrogenase (Acad or ACAD) gene family of enzymes includes very-long-chain (VLCAD), medium-chain (MCAD), and short-chain (SCAD) acyl-CoA dehydrogenases. The short-chain dehydrogenases/reductases family (SDR) constitute a large and diverse family of enzymes of ancient origin. Several of its members play an important role in human physiology and disease, especially in the metabolism of steroid substrates (e.g., prostaglandins, estrogens, androgens, and corticosteroids). Their involvement in common human disorders such as endocrine-related cancer, osteoporosis, and Alzheimer disease makes them an important candidate for drug targets.

As one of the first members of this family to be characterized was Drosophila alcohol dehydrogenase, and the family of this protein and related homologues are called 'insect-type', or 'short-chain' alcohol dehydrogenases ('adh-short'). A member of this protein family is the annotated protein product of Drosophila gene with GadFly Accession Number CG3842 containing the adh-short motif as major part of the protein (e.g., from amino acid 73 to amino acid 328 in the protein of 406 amino acids length). In humans, three proteins containing the adh-short motif were identified in this invention (see EXAMPLES) as homologs to the Drosophila CG3842 encoded protein. These proteins are CGI-82 (GenBank Accession Number NP_057110), PAN2 (GenBank Accession Number NP_065956), and the unnamed protein XP_085058 (GenBank Accession Number XP_085058).

The human CGI-82 gene was identified recently by comparative genomics (Lai et al., 2000, Genome Res 10(5):703-713). CGI-82 (PSDR1) is a member of the family of is a short-chain dehydrogenase/reductase enzymes (prostate short-chain dehydrogenase/reductase 1, PSDR1). The protein which is highly expressed in the prostate gland has been suggested to be

involved in the androgen receptor-regulated gene network of the human prostate. Genes regulated by androgenic hormones are of critical importance for the normal physiological function of the human prostate gland, and they contribute to the development and progression of prostate carcinoma. (Lin et al., 2001, Cancer Res 61(4):1611-1618).

The human PAN2 protein has been submitted to the NCBI Genbank recently (GenBank Accession Number NP_065965; submitted February 10, 2002 by Brereton et al.). PAN2 has been described as member of the SCAD superfamily.

So far, it has not been described that CG3842 encoded protein and closely related proteins, particularly SCAD proteins, such as human proteins CGI-82 (PSDR1), PAN2, and XP_85058 are involved in the regulation of energy homeostasis and body-weight regulation and related disorders, and thus, no functions in metabolic diseases and other diseases as listed above have been discussed. In this invention we demonstrate that the correct gene dose of CG3842 is essential for maintenance of energy homeostasis. A genetic screen was used to identify that mutation of a CG3842 homologous gene causes obesity, reflected by a significant increase of triglyceride content, the major energy storage substance.

Polynucleotides encoding a protein with homologies to CG3842, particularly a SCAD protien, are suitable to investigate diseases and disorders as described above. Further new compositions useful in diagnosis, treatment, and prognosis of diseases and disorders as described above are provided.

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described as these may vary. It is also to be understood that the terminology used

herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims. Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure.

The present invention discloses that CG3842 homologous proteins are regulating the energy homeostasis and fat metabolism especially the metabolism and storage of triglycerides, and polynucleotides, which identify and encode the proteins disclosed in this invention. The invention also relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides of the invention. The invention also relates to the use of these sequences in the diagnosis, study, prevention, and treatment of diseases and disorders, for example, but not limited to, metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea.

CG3842 homologous proteins and nucleic acid molecules coding therefor are obtainable from insect or vertebrate species, e.g. mammals or birds. Particularly preferred are human homologous nucleic acids, particularly

nucleic acids encoding a human PAN2 protein (GenBank Accession Number NP_065956 for the protein, NM_020905 for the cDNA), a human CGI-82 protein (GenBank Accession Number NP_057110 for the protein, NM_016026 for the cDNA), or an unnamed protein (GenBank Accession Number XP_085058 for the protein and GenBank Accession Number XM_085058 for the cDNA).

The invention particularly relates to a nucleic acid molecule encoding a polypeptide contributing to regulating the energy homeostasis and the metabolism of triglycerides, wherein said nucleic acid molecule comprises

(a) the nucleotide sequence of PAN2 (GenBank Accession Number NM_020905), human CGI-82 (GenBank Accession Number NM_016026), or a nucleotide sequence encoding an unnamed protein (GenBank Accession Number XM_085058), or GadFly Accession Number CG3842 and/or a sequence complementary thereto,

(b) a nucleotide sequence which hybridizes at 50°C in a solution containing 1 x SSC and 0.1% SDS to a sequence of (a),

(c) a sequence corresponding to the sequences of (a) or (b) within the degeneration of the genetic code,

(d) a sequence which encodes a polypeptide which is at least 85%, preferably at least 90%, more preferably at least 95%, more preferably at least 98% and up to 99,6% identical to the amino acid sequence of CG3842

(e) a sequence encoding a CG3842 homologous protein, preferably a human CG3842 homologous protein PAN2 (GenBank Accession Number NP_065956), human CGI-82 protein (GenBank Accession Number NP_057110), and unnamed protein with GenBank Accession Number XP_085058), and/or a sequence complementary thereto,

(e) a sequence which differs from the nucleic acid molecule of (a) to (d) by mutation and wherein said mutation causes an

alteration, deletion, duplication and/or premature stop in the encoded polypeptide or

(f) a partial sequence of any of the nucleotide sequences of (a) to (e) having a length of at least 15 bases, preferably at least 20 bases, more preferably at least 25 bases and most preferably at least 50 bases.

The invention is based on the finding that CG3842 homologous proteins, particularly proteins of the SCAD family, (herein referred to as CG3842) and the polynucleotides encoding these are involved in the regulation of triglyceride storage and therefore energy homeostasis. The invention describes the use of these compositions for the diagnosis, study, prevention, or treatment of diseases and disorders related thereto, including metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea.

Accordingly, the present invention relates to genes with novel functions in body-weight regulation, energy homeostasis, metabolism, and obesity. To find genes with novel functions in energy homeostasis, metabolism, and obesity, a functional genetic screen was performed with the model organism Drosophila melanogaster (Meigen). One resource for screening was a proprietary Drosophila melanogaster stock collection of PX-lines. The P-vector of this collection has Gal4-UAS-binding sites fused to a basal promoter that can transcribe adjacent genomic Drosophila sequences upon binding of Gal4 to UAS-sites. This enables the PX-line collection for overexpression of endogenous flanking gene sequences. In addition, without activation of the UAS-sites, integration of the EP-element into the gene is likely to cause a reduction of gene activity, and allows determining its function by evaluating the loss-of-function phenotype.

Triglycerides are the most efficient storage for energy in cells. In order to isolate genes with a function in energy homeostasis, several thousand proprietary PX-lines were tested for their triglyceride content after a prolonged feeding period (see Examples for more detail). Lines with significantly changed triglyceride content were selected as positive candidates for further analysis.

Obese people mainly show a significant increase in the content of triglycerides. In this invention, the content of triglycerides of a pool of flies with the same genotype after feeding for six days was analyzed using a triglyceride assay, as, for example, but not for limiting the scope of the invention, is described below in the examples section.

Flies homozygous for the integration of vectors for Drosophila line PX2287.1 were analyzed in an assay measuring the triglyceride contents of these flies, illustrated in more detail in the EXAMPLES. The result of the triglyceride content analysis is shown in FIGURE 1. The average triglyceride level of the fly collection in which the PX2287.1 line was found is shown as 100% in FIGURE 1 (First column, "TG010419, n=60"). The average increase of triglyceride content of the homozygous viable Drosophila line PX2287.1 is 80% (see FIGURE 1, second column, line "2287.1"). It was found in this invention that homozygous PX2287.1 flies have a significant higher triglyceride content than the control flies tested.

The increase of triglyceride content due to the loss of a gene function suggests gene activities in energy homeostasis in a dose dependent manner that controls the amount of energy stored as triglycerides.

Nucleic acids encoding the CG3842 protein of the present invention were identified using a plasmid-rescue technique. Genomic DNA sequences were isolated that are localized adjacent to the EP vector (herein PX2287.1) integration. Using those isolated genomic sequences public databases like

Berkeley Drosophila Genome Project (GadFly) or GenBank (NCBI) were screened thereby confirming the homozygous viable integration site of the PX2287.1 vector 542 base pairs downstream of the coding seqence of a gene, identified as Berkeley Drosophila Genome Project Accession Nr. CG3842 (FIGURE 2). FIGURE 2 shows the molecular organization of this gene locus. In FIGURE 2, genomic DNA sequence is represented by the assembly as a black thin line in the middle (numbers represent the length in basepairs of the genomic DNA) that includes the integration sites of vector for line PX2287.1. Transcribed DNA sequences (ESTs) and predicted exons are shown as bars on the two sides (sense and antisense strand). EST (expressed sequence tag) clones are represented as light grey bars, partly linked by thin lines, on the two outer sides. Predicted genes (as predicted by the Berkeley Drosophila Genome Project, GadFly) are represented by dark grey bars, linked by thin lines. Predicted exons of the Drosophila cDNA (annotated by Berkeley Drosophila Genome Project GadFly) are shown as dark grey bars and predicted introns as light grey lines.

The sequence of this invention encodes for a gene that is predicted by GadFly sequence analysis programs as Accession Number CG3842. Public DNA sequence databases (for example, NCBI GenBank) were screened thereby identifying the integration sites of line PX2287.1, causing an increase of triglyceride content. PX2287.1 is integrated 542 base pairs downstream of the coding seqence of a gene, identified as Berkeley Drosophila Genome Project Accession Nr. CG3842 (the site of integration is shown as vertical dotted line). Therefore, expression of the cDNA encoding Accession Number CG3842 could be effected by homozygous viable integration of vectors of line PX2287.1, leading to increase of the energy storage triglycerides.

The present invention is further describing a polypeptide comprising the amino acid sequence of CG3842. A comparison (Clustal X (1.81) analysis) between the CG3842 proteins of different species (human and Drosophila)

was conducted. Based upon homology, CG3842 protein of the invention and each homologous protein or peptide may share at least some activity. No functional data described the regulation of body weight control and related metabolic diseases such as obesity are available in the prior art for the genes of the invention.

The invention also encompasses polynucleotides that encode CG3842 and homologous proteins. Accordingly, any nucleic acid sequence, which encodes the amino acid sequences of CG3842, can be used to generate recombinant molecules that express CG3842. In a particular embodiment, the invention encompasses the nucleic acid sequence encoding a Drosophila protein (GadFly Accession Number CG3842), a human PAN2 protein (GenBank Accession Number NP_065956 for the protein, NM_020905 for the cDNA), a human CGI-82 protein (GenBank Accession Number NP_057110 for the protein, NM_016026 for the cDNA), or an unnamed protein (GenBank Accession Number XP_085058 for the protein and GenBank Accession Number XM_085058 for the cDNA). It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding CG3842, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequences of naturally occurring CG3842, and all such variations are to be considered as being specifically disclosed. Although nucleotide sequences which encode CG3842 and its variants are preferably capable of hybridizing to the nucleotide sequences of the naturally occurring CG3842 under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CG3842 or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which

expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CG3842 and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequences. The invention also encompasses production of DNA sequences, or portions thereof, which encode CG3842 and its derivatives, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art at the time of the filing of this application. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CG3842 any portion thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequences, and in particular, those of the polynucleotides comprising the nucleic acid sequence encoding a Drosophila protein (GadFly Accession Number CG3842), a human PAN2 protein (GenBank Accession Number NP_065956 for the protein, NM_020905 for the cDNA), a human CGI-82 protein (GenBank Accession Number NP_057110 for the protein, NM_016026 for the cDNA), or an unnamed protein (GenBank Accession Number XP_085058 for the protein and GenBank Accession Number XM_085058 for the cDNA). Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex or probe, as taught in Wahl, G. M. and S. L. Berger (1987: Methods Enzymol. 152:399-407) and Kimmel, A. R. (1987; Methods Enzymol. 152:507-511), and may be used at a defined stringency. Preferably, hybridization under stringent conditions means that after washing for 1 h with 1 x SSC and 0.1% SDS at 50°C, preferably at 55°C, more preferably at 62°C and most preferably at 68°C, particularly for 1 h in 0.2 x SSC and 0.1% SDS at 50°C,

preferably at 55°C, more preferably at 62°C and most preferably at 68°C, a positive hybridization signal is observed. Altered nucleic acid sequences encoding CG3842, which are encompassed by the invention include deletions, insertions, or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent CG3842.

The encoded proteins may also contain deletions, insertions, or substitutions of amino acid residues, which produce a silent change and result in a functionally equivalent CG3842. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological activity of CG3842 is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged amino acids may include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; phenylalanine and tyrosine.

Also included within the scope of the present invention are alleles of the genes encoding CG3842. As used herein, an "allele" or "allelic sequence" is an alternative form of the gene, which may result from at least one mutation in the nucleic acid sequence. Alleles may result in altered mRNAs or polypeptides whose structures or function may or may not be altered. Any given gene may have none, one, or many allelic forms. Common mutational changes, which give rise to alleles, are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence. Methods for DNA sequencing, which are well known and generally available in the art may be used to practice any embodiments of the invention. The methods may employ such

enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE DNA Polymerase (US Biochemical Corp, Cleveland Ohio), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, Ill.), or combinations of recombinant polymerases and proof-reading exonucleases such as the ELONGASE Amplification System (GIBCO/BRL, Gaithersburg, Md.). Preferably, the process is automated with machines such as the Hamilton MICROLAB 2200 (Hamilton, Reno Nev.), Peltier thermal cycler (PTC200; MJ Research, Watertown, Mass.) and the ABI 377 DNA sequencers (Perkin Elmer).

The nucleic acid sequences encoding CG3842 may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed, "restriction-site" PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). In particular, genomic DNA is first amplified in the presence of primer to linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase. Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). The primers may be designed using OLIGO 4.06 primer analysis software (National Biosciences Inc., Plymouth, Minn.), or another appropriate program, to 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68-72°C. The method uses several restriction enzymes to generate suitable fragments. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations also are used to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before performing PCR.

Another method, which may be used to retrieve unknown sequences is that of Parker, J. D. et al. (1991; Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries to walk in genomic DNA (Clontech, Palo Alto, Calif.). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences, which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into the 5' and 3' non-transcribed regulatory regions. Capillary electrophoresis systems, which are commercially available, may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled devise camera. Output/light intensity may be converted to electrical signal using appropriate software (e.g. GENOTYPER and SEQUENCE NAVIGATOR, Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially

preferable for the sequencing of small pieces of DNA, which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CG3842, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct expression of CG3842 in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences, which encode substantially the same, or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express CG3842. As will be understood by those of skill in the art, it may be advantageous to produce CG3842 -encoding and -encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence. The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter CG3842 encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CG3842 may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of CG3842 activities, it may be useful to encode

chimerical CG3842 proteins that can be recognized by commercially available antibodies. A fusion protein may also be engineered to contain a cleavage site located between the CG3842 encoding sequence and the heterologous protein sequences, so that CG3842 may be cleaved and purified away from the heterologous moiety. In another embodiment, sequences encoding CG3842 may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M. H. et al. (1980) Nucl. Acids Res. Symp. Ser. 7:215-223, Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 7:225-232). Alternatively, the proteins themselves may be produced using chemical methods to synthesize the amino acid sequence of CG3842, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J. Y. et al. (1995) Science 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin Elmer). The newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co., New York, N.Y.). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, supra). Additionally, the amino acid sequences of CG3842, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active CG3842, the nucleotide sequences encoding CG3842 functional equivalents, may be inserted into appropriate expression vectors, i.e., a vector, which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods, which are well known to those skilled in the art, may be used to construct expression vectors containing sequences encoding CG3842 and appropriate transcriptional and translational control elements. These

methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y.

A variety of expression vector/host systems may be utilized to contain and express sequences encoding CG3842. These include, but are not limited to, micro-organisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or PBR322 plasmids); or animal cell systems. The "control elements" or "regulatory sequences" are those non-translated regions of the vector-enhancers, promoters, 5' and 3' untranslated regions which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the BLUESCRIPT phagemid (Stratagene, LaJolla, Calif.) or PSPORT1 plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters and enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes) or from plant viruses (e.g., viral promoters and leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequences encoding CG3842,

vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for CG3842. For example, when large quantities of CG3842 are needed for the induction of antibodies, vectors, which direct high level expression of fusion proteins that are readily purified, may be used. Such vectors include, but are not limited to, the multifunctional E. coli cloning and expression vectors such as the BLUESCRIPT phagemid (Stratagene), in which the sequence encoding CG3842 may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of ß-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S. M. Schuster (1989) J. Biol. Chem. 264:5503-5509); and the like. PGEX vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with Glutathione S-Transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will. In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al., (supra) and Grantet al. (1987) Methods Enzymol. 153:516-544.

In cases where plant expression vectors are used, the expression of sequences encoding CG3842 may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or

heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp. 191-196).

An insect system may also be used to express CG3842. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding CG3842 may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and place under control of the polyhedrin promoter. Successful insertions of CG3842 will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells of Trichoplusia larvae in which CG3842 may be expressed (Engelhard, E. K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CG3842 may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain viable viruses that are capable of expressing CG3842 in infected host cells (Logan, J. and Shenk, T. (1984) Proc. Natl. Acad. Sci. 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CG3842. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding CG3842, its initiation codons, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, HEK293, and WI38, which have specific cellular machinery and characteriztic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines that stably express CG3842 may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a

selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells, which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type. Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1980) Cell 22:817-23) genes, which can be employed in tk-or aprt-,cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman, S. C. and R. C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, ß- glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C. A. et al. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequences encoding CG3842 are

inserted within a marker gene sequence, recombinant cells containing sequences encoding CG3842 can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with sequences encoding CG3842 under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well. Alternatively, host cells, which contain the nucleic acid sequences encoding CG3842 and express CG3842, may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA, or DNA-RNA hybridization and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

The presence of polynucleotide sequences encoding CG3842 can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or portions or fragments of polynucleotides specific for CG3842. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding CG3842 to detect transformants containing DNA or RNA encoding CG3842. As used herein "oligonucleotides" or "oligomers" refer to a nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides, which can be used as a probe or amplimer.

A variety of protocols for detecting and measuring the expression of CG3842, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CG3842 is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton, R.

et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, Minn.) and Maddox, D. E. et al. (1983; J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CG3842 include oligo-labeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide.

Alternatively, the sequences encoding CG3842, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, Mich.); Promega (Madison Wis.); and U.S. Biochemical Corp., (Cleveland, Ohio).

Suitable reporter molecules or labels, which may be used, include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, co-factors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CG3842 may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CG3842 may be designed to contain signal sequences, which direct secretion of

CG3842 through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding CG3842 to nucleotide sequence encoding a polypeptide domain, which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAG extension/affinity purification system (Immunex Corp., Seattle, Wash.) The inclusion of cleavable linker sequences such as those specific for Factor XA or Enterokinase (Invitrogen, San Diego, Calif.) between the purification domain and CG3842 may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing CG3842 and a nucleic acid encoding 6 histidine residues preceding a Thioredoxine or an Enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography as described in Porath, J. et al. (1992, Prot. Exp. Purif. 3: 263-281)) while the Enterokinase cleavage site provides a means for purifying CG3842 from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D. J. et al. (1993; DNA Cell Biol. 12:441-453). In addition to recombinant production, fragments of CG3842 may be produced by direct peptide synthesis using solid-phase techniques (Merrifield J. (1963) J. Am. Chem. Soc. 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A peptide synthesizer (Perkin Elmer). Various fragments of CG3842 may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

Diagnostics and Therapeutics

The data disclosed in this invention show that the nucleic acids and proteins of the invention are useful in diagnostic and therapeutic applications implicated, for example but not limited to, in metabolic disorders such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea. Hence, diagnostic and therapeutic uses for the CG3842 nucleic acids and proteins of the invention are, for example but not limited to, the following: (i) protein therapeutic, (ii) small molecule drug target, (iii) antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) diagnostic and/or prognostic marker, (v) gene therapy (gene delivery/gene ablation), (vi) research tools, and (vii) tissue regeneration in vitro and in vivo (regeneration for all these tissues and cell types composing these tissues and cell types derived from these tissues).

The nucleic acids and proteins of the invention are useful in diagnostic and therapeutic applications implicated in various applications as described below. For example, but not limited to, cDNAs encoding the CG3842 proteins of the invention and particularly their human homologues may be useful in gene therapy, and the CG3842 proteins of the invention and particularly their human homologues may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from, for example, but not limited to, in metabolic disorders as described above.

The novel nucleic acid encoding the CG3842 protein of the invention, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acids or the proteins are to

be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

For example, in one aspect, antibodies that are specific for CG3842 may be used directly as an antagonist, or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express CG3842. The antibodies may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimerical, single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with CG3842 any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminium hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvants used in human, BCG (Bacille Calmette-Guerin) and Corynebacterium parvum are especially preferable. It is preferred that the peptides, fragments, or oligopeptides used to induce antibodies to CG3842 have an amino acid sequence consisting of at least five amino acids, and more preferably at least 10 amino acids. It is preferable that they are identical to a portion of the amino acid sequence of the natural protein, and they may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of CG3842 amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule.

Monoclonal antibodies to CG3842 may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Köhler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R. J. et al. Proc. Natl. Acad. Sci. 80:2026-2030; Cole, S. P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison, S. L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M. S. et al (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CG3842 - and -specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D. R. (1991) Proc. Natl. Acad. Sci. 88:11120-3). Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments, which contain specific binding sites for CG3842, may also be generated. For example, such fragments include, but are not limited to, the $F(ab')_2$ fragments which can be produced by Pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy

identification of monoclonal Fab fragments with the desired specificity (Huse, W. D. et al. (1989) Science 254:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding and immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CG3842 and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CG3842 epitopes is preferred, but a competitive binding assay may also be employed (Maddox, supra).

In another embodiment of the invention, the polynucleotides encoding CG3842, or any fragment thereof, or antisense molecules, may be used for therapeutic purposes. In one aspect, antisense to the polynucleotide encoding CG3842 may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding CG3842. Thus, antisense molecules may be used to modulate CG3842 activity, or to achieve regulation of gene function. Such technology is now well know in the art, and sense or antisense oligomers or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding CG3842. Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods, which are well known to those skilled in the art, can be used to construct recombinant vectors, which will express antisense molecules complementary to the polynucleotides of the gene encoding CG3842. These techniques are described both in Sambrook et al. (supra) and in Ausubel et al. (supra). Genes encoding CG3842 can be turned off by transforming a cell or tissue

with expression vectors which express high levels of polynucleotide or fragment thereof which encodes CG3842. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA, or PNA, to the control regions of the gene encoding CG3842, i.e., the promoters, enhancers, and introns. Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it cause inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J. E. et al. (1994) In; Huber, B. E. and B. I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y.). The antisense molecules may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyse the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples, which may be used, include engineered hammerhead motif ribozyme molecules that can be specifically and efficiently catalyse endonucleolytic cleavage of sequences encoding CG3842. Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule

for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Antisense molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding CG3842. Such DNA sequences may be incorporated into a variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize antisense RNA constitutively or inducibly can be introduced into cell lines, cells, or tissues. RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5′ and/or 3′ ends of the molecule or the use of phosphorothioate or 2′ O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and

clonally propagated for autologous transplant back into that same patient. Delivery by transfection and by liposome injections may be achieved using methods, which are well known in the art. Any of the therapeutic methods described above may be applied to any suitable subject including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of CG3842, antibodies to CG3842, mimetics, agonists, antagonists, or inhibitors of CG3842. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs or hormones. The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries, which facilitate processing of the active compounds into preparations which, can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, Pa.). Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such

carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including Arabic and tragacanth; and proteins such as gelatine and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate. Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum Arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coating for product identification or to characterize the quantity of active compound, i.e., dosage. Pharmaceutical preparations, which can be used orally, include push-fit capsules made of gelatine, as well as soft, sealed capsules made of gelatine and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances, which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Optionally, the suspension may also contain suitable stabilizers or agents who increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes. The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulphuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use. After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of CG3842,

such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art. For any compounds, the therapeutically effective does can be estimated initially either in cell culture assays, e.g., of preadipocyte cell lines, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. A therapeutically effective dose refers to that amount of active ingredient, for example CG3842 fragments thereof, antibodies of CG3842, condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions, which exhibit large therapeutic indices, are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage from employed, sensitivity of the patient, and the route of administration. The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors, which may be taken into account, include the severity of the disease state, general health of the subject, age, weight, and gender of

the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation. Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

In another embodiment, antibodies which specifically bind CG3842 may be used for the diagnosis of conditions or diseases characterized by or associated with over- or underexpression of CG3842, or in assays to monitor patients being treated with CG3842, agonists, antagonists or inhibitors. The antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for CG3842 include methods, which utilize the antibody and a label to detect CG3842 in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules which are known in the art may be used several of which are described above.

A variety of protocols including ELISA, RIA, and FACS for measuring CG3842 are known in the art and provide a basis for diagnosing altered or abnormal levels of CG3842 expression. Normal or standard values for CG3842 expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to CG3842 under conditions suitable for complex formation. The

amount of standard complex formation may be quantified by various methods, but preferably by photometry, means. Quantities of CG3842 expressed in control and disease, samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides specific for CG3842 may be used for diagnostic purposes. The polynucleotides, which may be used, include oligonucleotide sequences, antisense RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of CG3842 may be correlated with disease. The diagnostic assay may be used to distinguish between absence, presence, and excess expression of CG3842, and to monitor regulation of CG3842 levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CG3842 closely related molecules, may be used to identify nucleic acid sequences which encode CG3842. The specificity of the probe, whether it is made from a highly specific region, e.g., unique nucleotides in the 5′ regulatory region, or a less specific region, e.g., especially in the 3′ coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences encoding CG3842, alleles, or related sequences. Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the CG3842 encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence of the polynucleotide comprising the nucleic acid sequence of nucleic acids encoding a Drosophila protein (GadFly Accession Number CG3842), a human PAN2 protein (GenBank

Accession Number NP_065956 for the protein, NM_020905 for the cDNA), a human CGI-82 protein (GenBank Accession Number NP_057110 for the protein, NM_016026 for the cDNA), or an unnamed protein (GenBank Accession Number XP_085058 for the protein and GenBank Accession Number XM_085058 for the cDNA), or from a genomic sequence including promoter, enhancer elements, and introns of the naturally occurring CG3842. Means for producing specific hybridization probes for DNAs encoding CG3842 include the cloning of nucleic acid sequences encoding CG3842 derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as ^{32}P or ^{35}S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CG3842 may be used for the diagnosis of conditions or diseases, which are associated with expression of CG3842. Examples of such conditions or diseases include, but are not limited to, pancreatic diseases and disorders, including diabetes. Polynucleotide sequences specific for CG3842 may also be used to monitor the progress of patients receiving treatment for pancreatic diseases and disorders, including diabetes. The polynucleotide sequences specific for CG3842 may be used in Southern or Northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dip stick, pin, ELISA or chip assays utilizing fluids or tissues from patient biopsies to detect altered CG3842 expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences specific for CG3842 may be useful in assays that detect activation or induction of various metabolic

diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancers of the reproductive organs, and sleep apnea. The nucleotide sequences encoding CG3842 may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the biopsied or extracted sample is significantly altered from that of a comparable have hybridized with nucleotide sequences in the sample, and the presence of altered levels of nucleotide sequences encoding CG3842 in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of an individual patient.

In order to provide a basis for the diagnosis of disease associated with expression of CG3842, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence which encodes CG3842 or a fragment thereof, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease. Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that, which is observed in

the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancers of the reproductive organs, and sleep apnea the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the pancreatic diseases and disorders. Additional diagnostic uses for oligonucleotides designed from the sequences encoding CG3842 may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced from a recombinant source. Oligomers will preferably consist of two nucleotide sequences, one with sense orientation (5'.fwdarw.3') and another with antisense (3'.rarw.5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantification of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of CG3842 include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby, P. C. et al. (1993) J. Immunol. Methods, 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantification of multiple samples may be accelerated by running

the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantification.

In another embodiment of the invention, the nucleic acid CG3842 sequences, which encode CG3842, may also be used to generate hybridization probes, which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. Such techniques include FISH, FACS, or artificial chromosome constructions, such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price, C. M. (1993) Blood Rev. 7:127-134, and Trask, B. J. (1991) Trends Genet. 7:149-154. FISH (as described in Verma et al. (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York, N.Y.) may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of the gene encoding CG3842 on a physical chromosomal map and a specific disease, or predisposition to a specific disease, may help to delimit the region of DNA associated with that genetic disease.

The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier, or affected individuals. In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This

provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti, R. A. et al. (1988) Nature 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequences of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

In another embodiment of the invention, CG3842, their catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening libraries of compounds, e.g. peptides or low molecular weight organic compounds, in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between CG3842 and the agent tested, may be measured.

Another technique for drug screening, which may be used, provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, as applied to CG3842 large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with CG3842, or fragments thereof, and washed. Bound CG3842 are then detected by methods well known in the art. Purified CG3842 can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support. In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CG3842 specifically compete

with a test compound for binding CG3842. In this manner, the antibodies can be used to detect the presence of any peptide, which shares one or more antigenic determinants with CG3842. In additional embodiments, the nucleotide sequences which encode CG3842 may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The Figures show:

Figure 1 shows the increase of triglyceride content of PX2287.1 flies (referred to as "2287.1") caused by integration of the P-vector (in comparison to controls with integration of these vectors elsewhere in genome, referred to as "TG010419, n = 60").

Figure 2 shows the molecular organization of the mutated CG3842 gene locus.

Figure 3 shows the BLASTP search results for CG3842 (Query) with the best human homologous matches (Sbject).
Figure 3A shows the homology to human unnamed protein with GenBank Accession Number XP_085058.1.
Figure 3B shows the homology to human PAN2 protein (GenBank Accession Number NP_065956.1).
Figure 3C shows the homology to human CGI-82 protein (GenBank Accession Number NP_057110.1).

Figure 4 shows the Clustal X (1.81) multiple sequence alignment analysis containing protein sequences for human CGI-82 (Accession Number NP_057110), human XP_085058, Drosophila GadFly Accession Number CG3842, and human PAN2 (Accession Number NP_065956)

The examples illustrate the invention:

Example 1: Measurement of triglyceride content

Mutant flies are obtained from a proprietary fly mutation stock collection. The flies are grown under standard conditions known to those skilled in the art. In the course of the experiment, additional feedings with bakers yeast (Saccharomyces cerevisiae) are provided. The average increase of triglyceride content of Drosophila flies containing the transposon vector in the homozygous viable PX2287.1 integration was investigated in comparison to control flies (FIGURE 1). For determination of triglyceride content, flies were incubated for 5 min at 90°C in an aqueous buffer using a waterbath, followed by hot extraction. After another 5 min incubation at 90°C and mild centrifugation, the triglyceride content of the flies extract was determined using Sigma Triglyceride (INT 336-10 or -20) assay by measuring changes in the optical density according to the manufacturer's protocol. As a reference protein content of the same extract was measured using BIO-RAD DC Protein Assay according to the manufacturer's protocol. The assay was repeated three times. The average triglyceride level of all flies of the PX collection is shown as 100% in FIGURE 1. PX2287.1 homozygous flies show constantly a higher triglyceride content than the controls The average increase of triglyceride content of the homozygous viable Drosophila line PX2287.1 is 80% (column 2 in FIGURE 1). Therefore, the change of gene activity in the locus of the PX2287.1 integration on chromosome X where the EP-vector of PX2287.1 flies is homozygous viable integrated, is responsible for changes in the metabolism of the energy storage triglycerides.

Example 2: Identification of the genes

In FIGURE 2, genomic DNA is represented by the assembly as a thin black line in the middle (numbers represent the length in basepairs of the

genomic DNA) that includes the integration sites of vector for line PX2287.1. Transcribed DNA sequences (ESTs) and predicted exons are shown as bars on the two sides (sense and antisense strand). Predicted exons of the cDNA are shown as dark grey bars and introns as light grey lines. The sequence encodes for a gene that is predicted by GadFly sequence analysis programs as Accession Number CG3842. Using those isolated genomic sequences public databases like Berkeley Drosophila Genome Project (GadFly) were screened confirming the homozygous viable integration site of the PX2287.1 vector 542 basepairs downstream of the coding sequence of CG3842, causing an increase of triglyceride content (the site of integration is shown as vertical dotted line). Therefore, expression of the cDNA encoding Accession Number CG3842 could be effected by homozygous viable integration of vectors of line PX2287.1, leading to an increase of the energy storage triglycerides.

Example 3: Identification of humanCG3842 homologues

CG3842 homologous proteins and nucleic acid molecules coding therefore are obtainable from insect or vertebrate species, e.g. mammals or birds. The most similar human nucleic acid sequences and the proteins encoded thereby have been determined using the BLAST algorithm searching public GenBank databases (see FIGURE 3). The most homologous human proteins are PAN2 (GenBank Accession Number NM_020905; 59% homology; see FIGURE 3B), human CGI-82 protein (GenBank Accession Number NM_016026; 62% homology; see FIGURE 3C), and unnamed protein with GenBank Accession Number XM_085058 (64% homology; see FIGURE 3A).

The results of a comparison of the Drosophila adh-short domains (e.g. GadFly Accession Number CG3842, amino acids 73 to 328) with the adh-short domains of human PAN2 protein (GenBank Accession Number

NM_020905), human CGI-82 protein (GenBank Accession Number NM_016026), and unnamed protein with GenBank Accession Number XP_085058 in a pairwise alignment are shown in TABLE 1.

TABLE 1. Results of pairwise alignment of deduced amino acid sequences of Drosophila GadFly Accession Numbers CG3842 and closely related human proteins XP_085058, CGI-82 (PSDR1), and PAN2.

Human protein	Drosophila protein	Identities/Similarities
XP_085058	CG3842	56%/67%
CGI-82	CG3842	55%/66%
PAN2	CG3842	52%/63%

A ClustaW (1.81) multiple sequence alignment has been conducted among the adh-short domains of the proteins described in TABLE 1 above and is shown in FIGURE 4.

Claims

1. A pharmaceutical composition comprising a nucleic acid molecule of the short-chain dehydrogenase (SCAD) gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an antibody, an aptamer or another receptor recognizing a nucleic acid molecule of the SCAD gene family or a polypeptide encoded thereby together with pharmaceutically acceptable carriers, diluents and/or adjuvants.

2. The composition of claim 1, wherein the nucleic acid molecule is a vertebrate or insect SCAD nucleic acid, particularly a nucleic acid encoding a Drosophila protein (GadFly Accession Number CG3842), a human PAN2 protein (GenBank Accession Number NP_065956 for the protein, NM_020905 for the cDNA), a human CGI-82 protein (GenBank Accession Number NP_057110 for the protein, NM_016026 for the cDNA), or an unnamed protein (GenBank Accession Number XP_085058 for the protein and GenBank Accession Number XM_085058 for the cDNA), or a fragment there of or a variant thereof and/or a nucleic acid complementary thereto.

3. The composition of claim 1 or 2, wherein said nucleic acid molecule
   (a) hybridizes at 50°C in a solution containing 1 x SSC and 0.1% SDS to a nucleic acid molecule as defined in claim 2 and/or a nucleic acid molecule which is complementary thereto;
   (b) it is degenerate with respect to the nucleic acid molecule of (a)
   (c) encodes a polypeptide which is at least 85%, preferably at least 90%, more preferably at least 95%, more preferably at least 98% and up to 99,6% identical to a SCAD polypeptide as defined in claim 2;

(d) differs from the nucleic acid molecule of (a) to (c) by mutation and wherein said mutation causes an alteration, deletion, duplication or premature stop in the encoded polypeptide.

4. The composition of any one of claims 1-3, wherein the nucleic acid molecule is a DNA molecule, particularly a cDNA or a genomic DNA.

5. The composition of any one of claims 1-4, wherein said nucleic acid encodes a polypeptide contributing to regulating the energy homeostasis and/or the metabolism of triglycerides.

6. The composition of any one of claims 1-5, wherein said nucleic acid molecule is a recombinant nucleic acid molecule.

7. The composition of any one of claims 1-6, wherein the nucleic acid molecule is a vector, particularly an expression vector.

8. The composition of any one of claims 1-5, wherein the polypeptide is a recombinant polypeptide.

9. The composition of claim 8, wherein said recombinant polypeptide is a fusion polypeptide.

10. The composition of any one of claims 1-7, wherein said nucleic acid molecule is selected from hybridization probes, primers and anti-sense oligonucleotides.

11. The composition of any one of claims 1-10 which is a diagnostic composition.

12. The composition of any one of claims 1-10 which is a therapeutic composition.

13. The composition of any one of claims 1-12 for the manufacture of an agent for detecting and/or verifying, for the treatment, alleviation and/or prevention of an disorders, including metabolic diseases such as obesity and other body-weight regulation disorders as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea and others, in cells, cell masses, organs and/or subjects.

14. Use of a nucleic acid molecule of the SCAD gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an antibody, an aptamer or another receptor recognizing a nucleic acid molecule of the SCAD gene family or a polypeptide encoded thereby for controlling the function of a gene and/or a gene product which is influenced and/or modified by a SCAD homologous polypeptide.

15. Use of the nucleic acid molecule of the SCAD gene family or a polypeptide encoded thereby or a fragment or a variant of said nucleic acid molecule or said polypeptide or an antibody, an aptamer or another receptor recognizing a nucleic acid molecule of the SCAD gene family or a polypeptide encoded thereby for identifying substances capable of interacting with a SCAD homologous polypeptide.

16. A non-human transgenic animal exhibiting a modified expression of a SCAD homologous polypeptide.

17. The animal of claim 16, wherein the expression of the SCAD homologous polypeptide is increased and/or reduced.

18. A recombinant host cell exhibiting a modified expression of a SCAD homologous polypeptide.

19. The cell of claim 18 which is a human cell.

20. A method of identifying a (poly)peptide involved in the regulation of energy homeostasis and/or metabolism of triglycerides in a mammal comprising the steps of

(a) contacting a collection of (poly)peptides with a SCAD homologous polypeptide or a fragment thereof under conditions that allow binding of said (poly)peptides;

(b) removing (poly)peptides which do not bind and

(c) identifying (poly)peptides that bind to said SCAD homologous polypeptide.

21. A method of screening for an agent which modulates the interaction of a SCAD homologous polypeptide with a binding target/agent, comprising the steps of

(a) incubating a mixture comprising

(aa)a SCAD homologous polypeptide, or a fragment thereof;

(ab) a binding target/agent of said SCAD homologous polypeptide or fragment thereof; and

(ac) a candidate agent

under conditions whereby said SCAD polypeptide or fragment thereof specifically binds to said binding target/agent at a reference affinity;

(b) detecting the binding affinity of said SCAD polypeptide or fragment thereof to said binding target to determine an (candidate) agent-biased affinity; and

(c) determining a difference between (candidate) agent-biased affinity and the reference affinity.

22. A method of producing a composition comprising the (poly)peptide identified by the method of claim 20 or the agent identified by the method of claim 21 with a pharmaceutically acceptable carrier, diluent and/or adjuvant.

23. The method of claim 22 wherein said composition is a pharmaceutical composition for preventing, alleviating or treating of diseases and disorders, including metabolic diseases such as obesity and other body-weight regulation disorders as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apneaand other diseases and disorders.

24. Use of a (poly)peptide as identified by the method of claim 20 or of an agent as identified by the method of claim 21 for the preparation of a pharmaceutical composition for the treatment, alleviation and/or prevention of of diseases and disorders, including metabolic diseases such as obesity and other body-weight regulation disorders as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancer, e.g. cancers of the reproductive organs, and sleep apnea and other diseases and disorders.

25. Use of a nucleic acid molecule of the SCAD family or of a fragment thereof for the preparation of a non-human animal which over- or under-expresses the CG3842 gene product.

26. Kit comprising at least one of

(a) a SCAD nucleic acid molecule or a fragment thereof;

(b) a vector comprising the nucleic acid of (a);

(c) a host cell comprising the nucleic acid of (a) or the vector of (b);

(d) a polypeptide encoded by the nucleic acid of (a);

(e) a fusion polypeptide encoded by the nucleic acid of (a);

(f) an antibody, an aptamer or another receptor against the nucleic acid of (a) or the polypeptide of (d) or (e) and

(g) an anti-sense oligonucleotide of the nucleic acid of (a).

Abstract

The present invention discloses CG3842 or SCAD homologous proteins regulating the energy homeostasis and the metabolism of triglycerides, and polynucleotides, which identify and encode the proteins disclosed in this invention. The invention also relates to the use of these sequences in the diagnosis, study, prevention, and treatment of diseases and disorders, for example, but not limited to, metabolic diseases such as obesity as well as related disorders such as eating disorder, cachexia, diabetes mellitus, hypertension, coronary heart disease, hypercholesterolemia, dyslipidemia, osteoarthritis, gallstones, cancers of the reproductive organs, and sleep apnea.

mh 07.03.2002

FIGURE 1. Triglyceride content of a CG3842 mutant

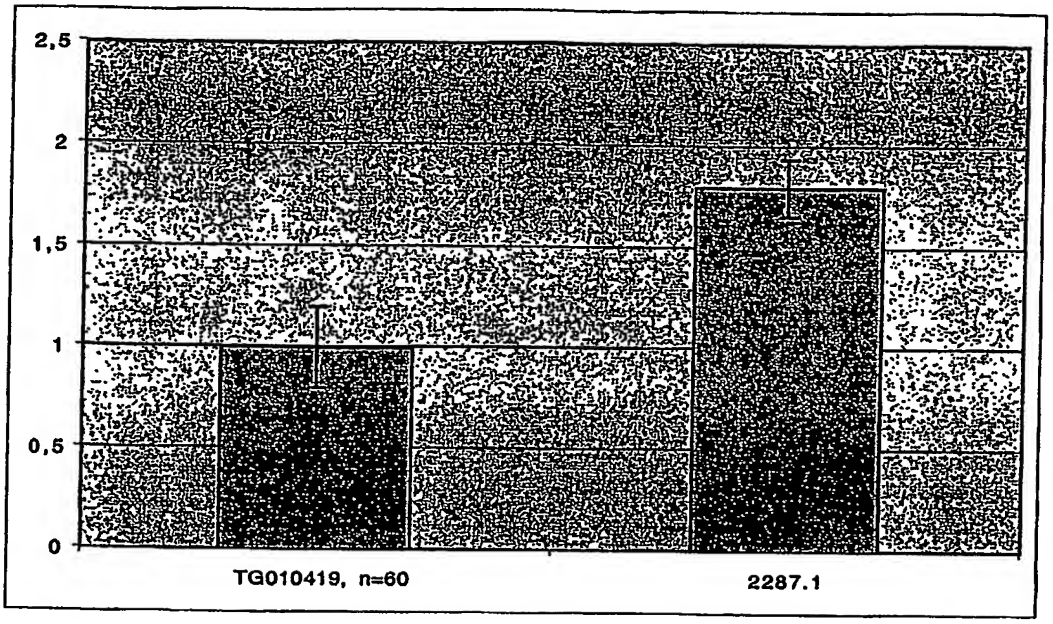

FIGURE 2. Molecular organisation of the gene CG3842 (GadFly Accession Number)

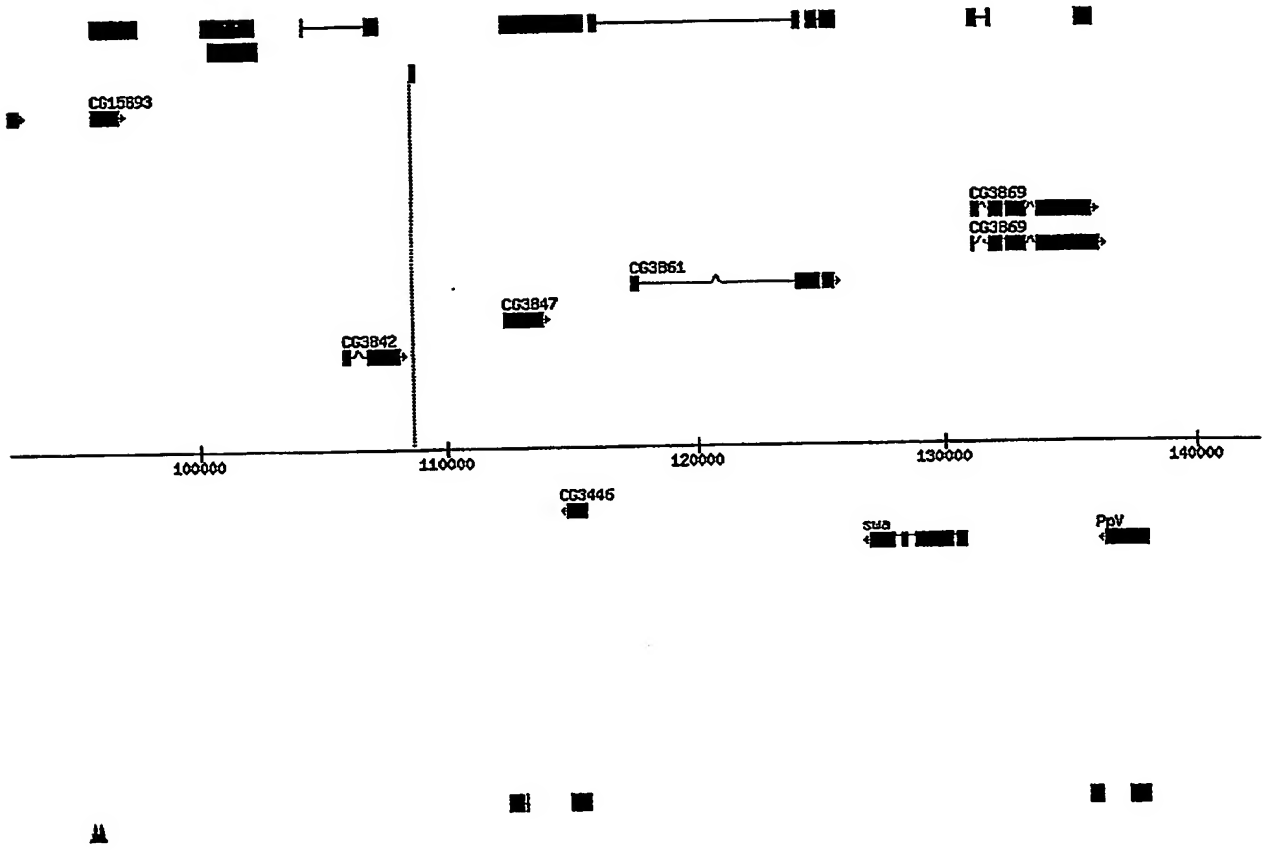

FIGURE 3. BLASTP RESULTS FOR CG3842

FIGURE 3A. Homology to human gene ref XM_085058, protein ref XP_085058.1

```
>ref|XP_085058.1| (XM_085058) similar to unnamed protein product [Homo sapiens]
 dbj|BAB70811.1| (AK054835) unnamed protein product [Homo sapiens]
Length = 316

Score =  266 bits (681), Expect = 2e-70
Identities = 163/317 (51%), Positives = 206/317 (64%), Gaps = 13/317 (4%)

Query: 45  LIVLGILL-----FMWL-----LRKCIQGPAYRKANRIDGKVVIVTGCNTGIGKETVLELAK 96
           L+ LG+L      F+++     +RK   G    R    ++ GKVV++TG NTGIGKET  ELA
Sbjct: 2   LVTLGLLTSFFSFLYMVAPSIRKFFAGGVCRTNVQLPGKVVVITGANTGIGKETARELAS 61

Query: 97  RGARVYMACRDPGRCEAARLDIMDRSRNQQLFNRTLDLGSLQSVRNFVERFKAEESRLDI 156
           RGARVY+ACRD  + E+A  +I    ++N Q+  R LDL    +S+R F E F AEE +L I
Sbjct: 62  RGARVYIACRDVLKGESAASEIRVDTKNSQVLVRKLDLSDTKSIRAFAEGFLAEEKQLHI 121

Query: 157 LINNAGVMACPRTLTADGFEQQFGVNHLGHFLLTNLLLDRLKHSSPSRIVVVSSAAHLFG 216
           LINNAGVM CP + TADGFE   GVNHLGHFLLT LLL+RLK S+P+R+V VSS AH  G
Sbjct: 122 LINNAGVMMCPYSKTADGFETHLGVNHLGHFLLTYLLLERLKVSAPARVVNVSSVAHHIG 181

Query: 217 RINREDLMSEKNYSKFFGAYSQSKLANILFTLKLSTILKDTGVTVNCCHPGVVRTEINRH 276
           +I   DL SEK YS+ F AY  SKLAN+LFT +L+  L+ TGVT    HPGVVR+E+ RH
Sbjct: 182 KIPFHDLQSEKRYSRGF-AYCHSKLANVLFTRELAKRLQGTGVTTYAVHPGVVRSELVRH 240

Query: 277 FSGPGWMKTALQKGSLYFFKTPKAGAQTQLRLALDPQLEGSTGGYYSDCMRWPLFPWVRN 336
            S    +   L +    F KT + GAQT L  AL   LE  +G Y+SDC R  + P  RN
Sbjct: 241 SS----LLCLLWRLFSPFVKTAREGAQTSLHCALAEGLEPLSGKYFSDCKRTWVSPRARN 296

Query: 337 MQTADWLWRESEKLLGL 353
           +TA+ LW  S +LLG+
Sbjct: 297 NKTAERLWNVSCELLGI 313
```

FIGURE 3B. Homology to human gene ref NM_020905, protein ref NP_065956.1

```
>ref|NP_065956.1| (NM_020905) PAN2 protein [Homo sapiens]
 gb|AAG12190.1|AF237952_1 (AF237952) PAN2 [Homo sapiens]
 gb|AAH09830.1|AAH09830 (BC009830) PAN2 protein [Homo sapiens]
 Length = 336

 Score =  254 bits (648), Expect = 1e-66
 Identities = 152/319 (47%), Positives = 191/319 (59%), Gaps = 20/319 (6%)

Query: 54  MWLLRKCIQGPAYRKANR------IDGKVVIVTGCNTGIGKETVLELAKRGARVYMACRD 107
           +WL  +   GP  ++  R       + GK V++TG N+G+G+ T  EL + GARV M CRD
Sbjct: 17  LWLAARRFVGPRVQRLRRGGDPGLMHGKTVLITGANSGLGRATAAELLRLGARVIMGCRD 76

Query: 108 PGRCEAARLDIMDRSRNQ-------------QLFNRTLDLGSLQSVRNFVERFKAEESRL 154
             R E A   +    R                +L  R LDL SL+SVR F +     EE RL
Sbjct: 77  RARAEEAAGQLRRELRQAAECGPEPGVSGVGELIVRELDLASLRSVRAFCQEMLQEEPRL 136

Query: 155 DILINNAGVMACPRTLTADGFEQQFGVNHLGHFLLTNLLLDRLKHSSPSRIVVVSSAAHL 214
           D+LINNAG+  CP   T DGFE QFGVNHLGHFLLTNLLL  LK S+PSRIVVVSS  +
Sbjct: 137 DVLINNAGIFQCPYMKTEDGFEMQFGVNHLGHFLLTNLLLGLLKSSAPSRIVVVSSKLYK 196

Query: 215 FGRINREDLMSEKNYSKFFGAYSQSKLANILFTLKLSTILKDTGVTVNCCHPGVVRTEIN 274
           +G IN +DL SE++Y+K F  YS+SKLANILFT +L+  L+ T VTVN  HPG+VRT +
Sbjct: 197 YGDINFDDLNSEQSYNKSF-CYSRSKLANILFTRELARRLEGTNVTVNVLHPGIVRTNLG 255

Query: 275 RHFSGPGWMKTALQKGSLYFFKTPKAGAQTQLRLALDPQLEGSTGGYYSDCMRWPLFPWV 334
           RH   P  +K      S  FFKTP  GAQT + LA  P++EG +G Y+ DC     L P
Sbjct: 256 RHIHIPLLVKPLFNLVSWAFFKTPVEGAQTSIYLASSPEVEGVSGRYFGDCKEEELLPKA 315

Query: 335 RNMQTADWLWRESEKLLGL 353
           +   A  LW  SE ++GL
Sbjct: 316 MDESVARKLWDISEVMVGL 334
```

FIGURE 3C. Homology to human gene ref NM_016026, protein ref NP_057110.1

```
>ref|NP_057110.1| (NM_016026) CGI-82 protein; likely ortholog of mouse cell line
          MC/9.IL4 derived transcript 1 [Homo sapiens]
 ref|XP_031073.1| (XM_031073) CGI-82 protein [Homo sapiens]
 gb|AAD34077.1|AF151840_1 (AF151840) CGI-82 protein [Homo sapiens]
 gb|AAH00112.1|AAH00112 (BC000112) CGI-82 protein [Homo sapiens]
 gb|AAK72049.1|AF395068_1 (AF395068) HCV core-binding protein HCBP12 [Homo sapiens]
 gb|AAH11727.1|AAH11727 (BC011727) Similar to CGI-82 protein [Homo sapiens]
 Length = 318

 Score =  250 bits (638), Expect = 2e-65
 Identities = 157/314 (50%), Positives = 196/314 (62%), Gaps = 7/314 (2%)

Query: 43  IFLIVLGILLFMWL--LRKCIQGPAYRKANRIDGKVVIVTGCNTGIGKETVLELAKRGAR 100
           + L++L  LL+M    +RK +          ++ GKVV+VTG NTGIGKET  ELA+RGAR
Sbjct: 8   LLLLLLPFLLYMAAPQIRKMLSSGVCTSTVQLPGKVVVVTGANTGIGKETAKELAQRGAR 67

Query: 101 VYMACRDPGRCEAARLDIMDRSRNQQLFNRTLDLGSLQSVRNFVERFKAEESRLDILINN 160
           VY+ACRD  + E     +I    + NQQ+  R LDL    +S+R F + F AEE  L +LINN
Sbjct: 68  VYLACRDVEKGELVAKEIQTTTGNQQVLVRKLDLSDTKSIRAFAKGFLAEEKHLHVLINN 127

Query: 161 AGVMACPRTLTADGFEQQFGVNHLGHFLLTNLLLDRLKHSSPSRIVVVSSAAHLFGRINR 220
           AGVM CP + TADGFE   GVNHLGHFLLT+LLL++LK S+PSRIV VSS AH  GRI+
Sbjct: 128 AGVMMCPYSKTADGFEMHIGVNHLGHFLLTHLLLEKLKESAPSRIVNVSSLAHHLGRIHF 187

Query: 221 EDLMSEKNYSKFFGAYSQSKLANILFTLKLSTILKDTGVTVNCCHPGVVRTEINRHFSGP 280
            +L  EK Y+    AY  SKLANILFT +L+  LK +GVT    HPG V++E+ RH S
Sbjct: 188 HNLQGEKFYNAGL-AYCHSKLANILFTQELARRLKGSGVTTYSVHPGTVQSELVRHSSFM 246

Query: 281 GWMKTALQKGSLYFFKTPKAGAQTQLRLALDPQLEGSTGGYYSDCMRWPLFPWVRNMQTA 340
           WM         +F KTP+ GAQT L  AL   LE  +G ++SDC    +    RN   A
Sbjct: 247 RWMWWLFS----FFIKTPQQGAQTSLHCALTEGLEILSGNHFSDCHVAWVSAQARNETIA 302

Query: 341 DWLWRESEKLLGLP 354
           LW  S  LLGLP
Sbjct: 303 RRLWDVSCDLLGLP 316
```

FIGURE 4. CLUSTAL X (1.81) multiple sequence alignment

```
CGI-82          PGKVVVVTGANTGIGKETAKELAQRGARVYLACRDVEKGELVAKEIQTTTGNQ-------
XP_085058       PGKVVVITGANTGIGKETARELASRGARVYIACRDVLKGESAASEIRVDTKNS-------
cg3842          DGKVVIVTGCNTGIGKETVLELAKRGARVYMACRDPGRCEAARLDIMDRSRNQ-------
PAN2            HGKTVLITGANSGLGRATAAELLRLGARVIMGCRDRARAEEAAGQLRRELRQAAECGPEP
                 *:. ::**.* *:*: *. *:    *. * :.***  : * .  ::     :     .

CGI-82          ------QVLVRKLDLSDTKSIRAFAKGFLAEEK-HLHVLINNAGVMMCPYS-KTADGFEM
XP_085058       ------QVLVRKLDLSDTKSIRAFAEGFLAEEK-QLHILINNAGVMMCPYS-KTADGFET
cg3842          ------QLFNRTLDLGSLQSVRNFVERFKAEES-RLDILINNAGVMACPRT-LTADGFEQ
PAN2            GVSGVGELIVRELDLASLRSVRAFCQEMLQEEP-RLDVLINNAGIFQCPYM-KTEDGFEM
                      ::: * *** .  *:* *   :  *:  :*.:******:: **    * **:*

CGI-82          HIGVNHLGHFLLTHLLLEKLKESAPSRIVNVSSLAHHLGRIHFHNLQGEKFYNAGL-AYC
XP_085058       HLGVNHLGHFLLTYLLLERLKVSAPARVVNVSSVAHHIGKIPFHDLQSEKRYSRGF-AYC
cg3842          QFGVNHLGHFLLTNLLLDRLKHSSPSRIVVVSSAAHLFGRINREDLMSEKNYSKFFGAYS
PAN2            QFGVNHLGHFLLTNLLLGLLKSSAPSRIVVVSSKLYKYGDINFDDLNSEQSYNKSF-CYS
                ::****:****** ***  ** ::*:*:* :**  :  * *   :* .*: *.  : .*.

CGI-82          HSKLANILFTQELARRLKGSGVTTYSVHPGTVQSELVRHSS----FMRWMWWLFS-----
XP_085058       HSKLANVLFTRELAKRLQGTGVTTYAVHPGVVRSELVRHSS----LLCLLWRLFS-----
cg3842          QSKLANILFTLKLSTILKDTGVTVNCCHPGVVRTEINRHFS----GPGWMKTALQ-K-GS
PAN2            RSKLANILFTRELARRLEGTNVTVNVLHPGIVRTNLGRHIH----IPLLVKPLFN--LVS
                :*****:*** :*:  *..:.**.   *** * ::: *:          .   :

CGI-82          -FFIKTPQQGAQTSLHCALTEGLEILSGNHFSDCHVA
XP_085058       -PFVKTAREGAQTSLHCALAEGLEPLSGKYFSDCKRT
cg3842          LYFFKTPKAGAQTQLRLALDPQLEGSTGGYYSDCMRW
PAN2            WAFFKTPVEGAQTSIYLASSPEVEGVSGRYFGDCKEE
                  :.**.  **** :  *    ::  :* ::.**
```